多维对数正态作答时间模型：对潜在加工速度多维性的探究*

詹沛达 [1]　Hong Jiao[2]　Kaiwen Man[3]

([1] 浙江师范大学教师教育学院心理学系，金华 321004)

([2]Measurement, Statistics, and Evaluation, Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland, United States)

([3]Educational Studies in Psychology, Research Methodology, and Counseling, The University of Alabama, Tuscaloosa, United States)

摘　要　在心理与教育测量中，潜在加工速度反映学生运用潜在能力解决问题的效率。为在多维测验中探究潜在加工速度的多维性并实现参数估计，本研究提出多维对数正态作答时间模型。实证数据分析及模拟研究结果表明：(1)潜在加工速度具有与潜在能力相匹配的多维结构；(2)新模型可精确估计个体水平的多维潜在加工速度及与作答时间有关的题目参数；(3)冗余指定潜在加工速度具有多维性带来的负面影响低于忽略其多维性所带来的。

关键词　题目作答时间; 多维潜在加工速度; 题目作答理论；计算机化测验; PISA

收稿日期: 2020-03-02

* 国家自然科学基金青年基金项目(31900795)资助。

通讯作者: 詹沛达, E-mail: pdzhan@gmail.com

1 引 言

近些年，随着计算机化测验的普及，对题目作答时间(response times, RT)及其他过程数据的采集已趋于常态化。例如，自 2012 年以来，国际学生能力评估项目(PISA)就开始采用计算机化测验采集学生的 RT 数据。已有研究表明，RT 数据作为传统作答精度数据外的一种补充，不仅能够提供学生在问题解决中的加工速度信息，在联合分析中还可以提高对潜在能力的测量精度(Bolsinova & Tijmstra, 2017; van der Linden, Klein Entink, & Fox, 2010; 詹沛达, 2019)。因此，近些年对 RT 数据的分析成为了国内外心理与教育测量领域的新热点之一。

研究者基于认知心理学理论和实验研究提出了多种 RT 模型(参见 De Boeck & Jeon, 2019; 郭磊，尚鹏丽，夏凌翔, 2017)。其中，速度-精度权衡(speed-accuracy trade-off; Luce, 1986)是一些早期 RT 模型所探讨的主要议题(例如, Ferrando & Lorenzo-Seva, 2007; Wang, T. & Hanson, 2005)，即对于特定的任务，被试的加工速度越快则其加工精度(或成功率)越低；反之，被试的加工速度越慢则其加工精度越高。然而，该权衡反映的是加工速度与加工精度在个体内(within-person)的关系(van der Linden, 2009)，无法通过横断研究/测验来评估(Curran & Bauer, 2011)。通常，对于一组固定的任务/题目，一旦被试的加工速度被固定，那么其加工精度也是固定的；因此，建议对加工速度和加工精度分别建模，而与之相对应的潜在加工速度和潜在能力之间的关系可以在更高的层次上建构(van der Linden, 2006; 2007; 2009)。当前，使用最多的是对数正态 RT 模型(lognormal RT model, LRTM) (van der Linden, 2006)，也有一些研究对该模型做了进一步拓广(例如，孟祥斌, 2016; Klein Entink, van der Linden, & Fox, 2009; Wang, C., Chang, & Douglas, 2013)。

为进一步探究潜在加工速度与潜在能力之间的关系，van der Linden (2007)提出了贝叶斯层级建模框架。该框架的基本逻辑是，在个体内，潜在加工速度对 RT 的影响和潜在能力对作答精度(response accuracy, RA)的影响是相互独立的；而在群体内(即个体间)，潜在加工速度与潜在能力之间具有相关关系。鉴于该框架的灵活性，通过替换不同的测量模型已形成多种联合模型(例如，詹沛达, 2019; Guo, Luo, & Yu, 2020; Lu, Wang, Zhang, & Tao, 2019; Man, Harring, Jiao, & Zhan, 2019; Wang, C. & Xu, 2015; Wang, Zhang, Douglas, & Culpepper, 2018; Zhan, Jiao, & Liao, 2018)。但目前，绝大多数联合模型都仅适用于单维测验，即使用单维题目作答理论(item response theory)模型来分析 RA 数据并使用单维 RT 模型来分析 RT 数据；而仅有的几个模型虽然关注到了潜在能力的多维性问题，但仍假设潜在加工速度是单维的，进而使用多维 IRT (multidimensional IRT, MIRT)模型分析 RA 数据并仍使用单维 RT 模型来分析 RT 数据(詹沛达, 2019; Man et al., 2019; Wang, Weiss, & Su, 2019; Zhan, Jiao et al., 2018)。导致该问题的主要原因是目前尚未有研究者关注到潜在加工速度可能存在多维性的问题，同时也缺少相应的分析模型。

在心理和教育测量中，关于潜在加工速度的一个恰当的概念是劳动的速度(speed of labor) (van der Linden, 2009)。因此，潜在加工速度可被定义为“解答某题目时所付出劳动与所花费时间的比例(a rate of

the amount of labor performed on the items with respect to time)”(van der Linden, 2011)。潜在加工速度反映了学生运用潜在能力(例如, 知识或技能)来解决问题的效率。针对同一道题目，学生消耗的作答时间越少表明其潜在加工速度越快，反映出学生运用该题目所需的知识或技能的效率越高。在多维测验中，由于潜在能力的多维性，潜在加工速度应该在特定的测验维度中与潜在能力一起讨论，即潜在加工速度也可能具有与潜在能力相匹配的多维结构。换句话说，被试在每个测验维度上的潜在加工速度与该维度所需的潜在能力相匹配。例如，被试在解码任务中的潜在加工速度与该任务所需的解码能力相匹配，而被试在编码任务中的潜在加工速度与该任务所需的编码能力相匹配。再比如，当非英语母语被试参加 GRE 学科测验(例如, 数学或英语文学)时，至少需要两个潜在能力，一个用于理解题目(例如, 英语阅读能力)，一个用于解决问题(例如, 学科能力)。这会涉及到对应的两个潜在加工速度，一个反映理解题目的速度，一个反映解决问题的速度。

对此，本研究假设：在多维测验中，潜在加工速度具有与潜在能力相匹配的多维结构。已有一些认知心理学证据可能支持该假设。首先，不同的大脑区域工作对应于不同的认知加工功能，适当的行为表现取决于特定大脑区域之间的相互作用(Horwitz et al., 1999; Mesulam, 1990)，这也是功能磁共振成像(fMRI)和脑电图(EEG)的基本逻辑。从概念上讲，不同认知任务所需的不同认知加工功能具有不同的认知加工速度。其次，与在实验心理学中用来记录反应时(reaction time)的简单刺激任务(例如, 数字广度任务[digit-span task]等其他不涉及特定陈述性和程序性知识的刺激任务)不同，心理和教育测量中的题目始终是对特定认知建构或能力的测查。因此，在心理和教育测量中观察到的 RT 应包括两个部分：用于加工所有信息的基本反应时和运用特定潜在能力所消耗的时间。鉴于题目水平的 RT 无法区分两者，所以必须将它们视为一个整体来看待。此时，我们可以使用“特定维度的加工时间(dimension-specific processing time)”来指代题目水平 RT，并使用“特定维度的加工速度(dimension-specific processing speed)”来指代多维潜在空间中特定维度中的加工速度。因此，与潜在能力一样，潜在加工速度的维度数也可由测验所包含的维度数来确定。

目前，尽管针对 RA 的 MIRT 模型已经得到较好的发展(Reckase, 2009)，但尚缺乏可分析多维潜在加工速度的多维 RT (multidimensional RT, MRT)模型。如上文所述，近期已有一些研究尝试使用 MIRT 模型来分析多维潜在能力，但仍使用 URT 模型来分析可能存在的多维潜在加工速度(詹沛达, 2019; Man et al., 2019; Wang, C. et al., 2019)。然而，由于缺少 MRT 模型，上述研究仅能估计学生的多个潜在能力和一个潜在加工速度。从逻辑上讲不同的潜在能力应与不同的潜在加工速度相匹配；因此，强制将多个潜在加工速度约束为一个变量的做法具有局限性，可能导致推论不准确。在多维测验中，尽管单维潜在加工速度可以被解释为被试的一般或高阶潜在加工速度，但实际上，我们仍渴望知道被试在每一个子维度上的潜在加工速度。因此，开发相应的 MRT 模型是有必要的。

为解决上述问题，本研究提出了多维对数正态 RT 模型(multidimensional LRTM, MLRTM)。该模型

可视为对单维对数正态 RT 模型 (unidimensional LRTM, ULRTM) (van der Linden, 2006)的拓广。首先，简单回顾了 ULRTM；其次，提出了 MLRTM；然后，对 2012 年 PISA 计算机化数学测验中 RT 数据进行了探索性因素分析以探究潜在加工速度的多维结构，使用新提出的模型对该数据做进一步分析，并与 ULRTM 进行对比，以展现新模型的实际可应用性和相对优势；随后，通过一则模拟研究来探究新模型的心理计量学性能；最后，总结了研究结果并讨论了未来的研究方向。

2 多维对数正态作答时间模型

2.1 模型建构

在介绍 MLRTM 前，我们先简单回顾下 ULRTM。设定 T_{ni} 为学生 n ($n = 1, ..., N$)对题目 i ($i = 1, ..., I$)的作答时间。则 ULRTM 可表示为

$$\log T_{ni} = \xi_i - \tau_n + \varepsilon_{ni},\ \varepsilon_{ni} \sim N(0,\ \omega_i^{-2}), \tag{1}$$

或

$$\log T_{ni} \sim N(\xi_i - \tau_n,\ \omega_i^{-2}). \tag{2}$$

其中，ξ_i 为题目时间强度参数，表示解答题目 i 所必需的时间；τ_n 是学生 n 的潜在加工速度，假定其满足 $\tau_n \sim N(0,\sigma_\tau^2)$；$\varepsilon_{ni}$ 为残差；ω_i 是残差的标准差的倒数，可以将其视为题目时间区分度参数。ULRTM 的基本假设之一是 $\log T_{ni}$ 在给定单维τ_n 时满足条件独立。

在心理与教育测量中，主要有两种多维测验类型：题目内(within-item)和题目间(between-item) (Adams, Wilson, & Wang, W.-C., 1997)。在题目间多维测验中，每个题目仅测量一个维度的潜在能力，但不同题目可能会测量不同维度的潜在能力；而在题目内多维测验中，一个题目可能同时测量多个维度的潜在能力。从理论上讲，题目间多维度是题目内多维度的一个特例，因此，本研究借鉴题目内多维度的表达式来建构 MLRTM。则 MLRTM 可表示为

$$\log T_{ni} = \xi_i - \sum_{k=1}^{K} \tau_{nk} q_{ik} + \varepsilon_{ni},\ \varepsilon_{ni} \sim N(0,\ \omega_i^{-2}), \tag{3}$$

或

$$\log T_{ni} \sim N(\xi_i - \sum_{k=1}^{K} \tau_{nk} q_{ik},\ \omega_i^{-2}), \tag{4}$$

其中，τ_{nk} 是学生 n 在维度 k ($k = 1, 2, ..., K$)上的潜在加工速度，反映了学生 n 运用第 k 维度潜在能力来解决问题的效率；$\boldsymbol{\tau}_n = (\tau_{n1}, ..., \tau_{nk}, ..., \tau_{nK})'$是遵循多元正态分布的多维潜在加工速度向量：$\boldsymbol{\tau}_n \sim N(\boldsymbol{\mu}_\tau, \boldsymbol{\Sigma}_\tau)$，其中均值向量$\boldsymbol{\mu}_\tau = (\mu_1, ..., \mu_k, ..., \mu_K)'$和方差-协方差矩阵$\boldsymbol{\Sigma}_\tau$，$\mu_k$ 是维度 k 上学生总体的平均加工速度。为使模型可识别，将$\boldsymbol{\mu}_\tau$设置为 **0** 向量。Q 矩阵(Tatsuoka, 1983)是一个 $I \times K$ 的验证性矩阵，其中 $q_{ik} = 1$ 表示题目 i 归属于维度 k，反之 $q_{ik} = 0$。对于题目间多维度，$\mathbf{q}_i$ 中只有一个元素等于 1；对于题目内多维度，

$\mathbf{q}_i$ 中有多个元素等于 1。其他参数与 ULRTM 中的参数相同。在 MLRTM 中，假定 $\log T_{ni}$ 在给定 $\boldsymbol{\tau}_n$ 的情况下满足条件独立。此外，若假定测验中所有题目仅考查同一个维度，则 MLRTM 等价于 ULRTM。

2.2 贝叶斯参数估计

本研究使用全贝叶斯马尔可夫链蒙特卡洛算法对 MLRTM 进行参数估计，并基于 MultiBUGS (version 1.0) (Goudie, Turner, De Angelis, & Thomas, 2017) 实现。感兴趣的读者可向通讯作者索取 MultiBUGS 代码，MLRTM 中各待估计参数的先验分布设定详见附录。

3. 实证数据分析

3.1 潜在加工速度多维结构的探索

如上文所述，本研究的基本假设是，在多维测验中，潜在加工速度具有与潜在能力相匹配的多维结构。为了探索潜在加工速度的多维性，并探究潜在加工速度的多维结构是否与潜在能力的多维结构相匹配，我们拟对一则 RT 实证数据进行探索性因素分析。

3.1.1 数据描述

本研究选用 2012 年 PISA 计算机化数学测验中的 RT 数据。该数据集最初由 Zhan, Jiao et al. (2018)使用。该数据包含 $N = 1581$ 名学生对 $I = 9$ 道题目的作答。原始 RT 数据均事先求取对数，并将所有 0 视为缺失数据。Zhan, Jiao et al. (2018)根据 2012 年 PISA 数学测评框架(OECD, 2013)设定了 Q 矩阵[1]，本研究选择了属于数学内容知识的三个维度，即变化和关系(θ_1)，空间和形状(θ_2)，以及不确定性和数据(θ_3)，见表 1。需要强调的是，该 Q 矩阵界定了题目和潜在能力之间关系，即该 Q 矩阵表达的是 RA 数据背后的潜在能力的多维结构。此时，若该 Q 矩阵与通过对 RT 数据进行探索性因素分析发现的潜在结构(即 RT 数据背后的潜在加工速度的结构)相匹配，就可说明潜在加工速度具有与潜在能力相匹配的多维结构。

表 1. 2012 年 PISA 计算机化数学测验的 Q 矩阵.

题目	θ_1	θ_2	θ_3
CM015Q02D	1		
CM015Q03D	1		
CM020Q01		1	
CM020Q02		1	
CM020Q03		1	
CM020Q04		1	
CM038Q03T			1
CM038Q05			1
CM038Q06			1

注：空白表示“0”.

[1] Q 矩阵本质上只是一个验证性矩阵，用于界定题目与潜在变量之间的关系，其使用范围并不局限在认知诊断领域，且其中的潜在变量也并不限定于知识、技能等细颗粒属性。

3.1.2 探索性因素分析

本研究使用 Mplus (version 8.1) (Muthén & Muthén, 2019)进行探索性因素分析。Mplus 默认使用验证性因素分析框架下的探索性因素分析，本研究将保留因素数量设为从 1 到 5。根据模型-数据拟合指标(例如, AIC 和 BIC)来确定因素数量以及相应的潜在结构。理论上，多个维度之间应该存在相关，因此使用斜交旋转。其他均采用默认设置。

表 2 给出了探索性因素分析的模型-数据拟合指标。前人研究表明 TLI > 0.95，CFI > 0.95，SRMR < 0.08，RMSEA < 0.05 意味着良好的模型-数据拟合(Hu & Bentler, 1999; Steiger, 1990)。综合各个指标，可认为三因素模型比其他模型更适合该数据，表明 RT 数据背后具有三维潜在结构。

表 3 给出了三因素模型的旋转因素载荷矩阵。可发现，该因素载荷矩阵与表 1 中的 Q 矩阵相比，仅题目 CM038Q03T 存在差异，且 CM038Q03T 在因素 3 上的载荷为 0.300 ($p < 0.05$)。因此，可以说由理论构建的潜在能力的多维结构(即 Q 矩阵)与对 RT 数据进行探索性因素分析发现的潜在结构是相匹配的。该结果支持了本研究的核心假设，即在多维测验中，潜在加工速度具有与潜在能力相匹配的多维结构。因此，后续研究可直接使用表 1 中的 Q 矩阵来表达 RA 和 RT 数据背后一致的多维潜在结构。当然，由于探索性因素分析本身的限制，我们无法获得每位学生的潜在加工速度估计值以及每道题目的题目参数。因此，有必要进一步利用本研究提出的 MLRTM 进行数据分析。

表 2. 2012 年 PISA 计算机化数学测验数据的探索性因素分析中的数据-模型拟合指标.

Model	χ^2	*df*	TLI	CFI	AIC	BIC	SRMR	RMSEA [90% CI]
1-factor	462.79**	27	0.896	0.922	24592.15	24737.03	0.045	0.101 [0.093, 0.109]
2-factor	225.49**	19	0.930	0.963	24370.85	24558.65	0.032	0.083 [0.073, 0.093]
3-factor	32.66**	12	0.989	0.996	24192.02	24417.38	0.010	0.033 [0.020, 0.047]
4-factor	5.56	6	1.000	1.000	24176.92	24434.48	0.004	0.000 [0.000, 0.031]
5-factor	0.09	1	1.006	1.000	24181.44	24465.83	0.000	0.000 [0.000, 0.045]

注：** = $p < .01$; χ^2 = 卡方; *df* = 自由度; TLI = Tucker-Lewis index; CFI = comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; 90% CI = 90%置信区间.

表 3. 三因素模型的旋转因素载荷矩阵.

题目	因素 1	因素 2	因素 3
CM015Q02D	0.695*		
CM015Q03D	0.609*		
CM020Q01		0.565*	
CM020Q02		0.801*	
CM020Q03		0.642*	
CM020Q04		0.943*	
CM038Q03T		0.502*	
CM038Q05			0.985*
CM038Q06			0.621*

注：* = $p < .05$; 未呈现因素载荷.4 以下的值.

3.2 采用多维对数正态作答时间模型进行分析

3.2.1 分析

为实现对 RT 数据的深入分析，本研究同时使用 ULRTM 和 MLRTM 分析该数据。探索性因素分析结果表明表 1 中的 Q 矩阵适用于描述题目和潜在加工速度之间的关系。在贝叶斯 MCMC 估计中设定 2 条马尔可夫链，每条链包含 5000 次迭代(其中前 2000 次做 burn-in)，最后保留两条链剩余的共 6000 次迭代进行参数估计推断。使用 MC_error 指标进行参数估计收敛检验(Ntzoufras, 2009)，本研究所有参数的 MC_error 均小于 0.05，表示参数估计已收敛。

本研究使用 DIC 和 WAIC (Gelman et al., 2014, Chapter 7)作为模型-数据相对拟合指标进行模型选择。使用后验预测模型检验(posterior predictive model checking, PPMC)来评估模型-数据绝对拟合，其中后验预测概率(posterior predictive probability, *ppp*)接近 0.5 表明模型与数据拟合。对 PPMC 而言选取一个合适的差异测度的必要的，本研究选用被试 n 和题目 i 的标准化误差函数之和作为差异测量(Fox & Marianti, 2017)来评估 RT 模型的整体拟合情况：

$$D(\log \mathbf{T};\boldsymbol{\upsilon}) = D(\log T_{ni};\xi_i,\boldsymbol{\tau}_n,\omega_i) = \sum_{n=1}^{N}\sum_{i=1}^{I}\left(\omega_i(\log T_{ni} - (\xi_i - \sum_{k=1}^{K} q_{ik}\tau_{nk}))\right)^2 .$$

3.2.2 结果

表 4 呈现了模型-数据拟合指标。其中，ULRTM 和 MLRTM 的 *ppp* 值分别为 0.597 和 0.633，表明这两个模型均拟合该数据。进一步，由–2LL、DIC 和 WAIC 指标均表示 MLRTM 对该数据的拟合程度更高，说明在多维测验中考虑潜在加工速度的多维性是更合适的。

表 5 呈现了方差-协方差矩阵估计值。三个潜在加工速度之间的相关系数范围为 0.751 到 0.855，表明这三个潜在加工速度为中等偏高程度相关，即三者之间有较高一致性但仍清晰可分。主要原因是三者都归属于数学内容知识这一更高阶的维度。另外，ULRTM 中单维潜在加工速度的方差估计值为 0.216 (95%CI = [0.197, 0.231]) ，不仅无法区分不同维度上的潜在加工速度，还低估了维度 1 (变化和关系)和维度 3 (不确定性和数据)上所有被试的潜在加工速度之间的差异性(即方差被低估)。

图 1 呈现了前 20 名被试的潜在加工速度估计值。根据 MLRTM 的估计结果，每个被试在 3 个维度上的潜在加工速度都是不同的，甚至有一些被试(例如, 被试 2、6、7、12、15)在 3 个维度上的潜在加工速度估计值的正负号都不同。此时，若使用 ULRTM 中的单维估计值作为被试的反馈信息(甚至基于此给被试贴上诸如“急先锋”或“慢郎中”的标签)势必过于笼统，无法体现出被试在不同维度上潜在加工速度之间的差异。

表 6 呈现了题目参数估计值。对题目时间强度参数而言，两模型的参数估计结果基本一致，表明考虑潜在加工速度的多维性并不影响题目时间强度参数的估计。与之相比，MLRTM 对题目时间区分度参数的估计值略大于 ULRTM 的，即 ULRTM 会低估 log RT 的峰度值。

表 4. 2012 年 PISA 计算机化数学测验数据分析中模型-数据拟合指标

分析模型	–2LL	DIC	WAIC	*ppp*

MLRTM	19305	22505	22055	0.633
ULRTM	21310	22890	22770	0.597

注：ULRTM = 单维对数正态作答时间模型; MLRTM = 多维对数正态作答时间模型; –2LL = –2 log likelihood; DIC = deviance information criterion; WAIC = widely available information criterion; *ppp* = 后验预测概率.

表 5. 2012 年 PISA 计算机化数学测验数据分析中多维潜在加工速度的方差-协方差矩阵估计值

$\mathbf{\Sigma}_\tau$	τ_1	τ_2	τ_3
τ_1	0.301 (0.016) [0.270, 0.334]	0.751	0.767
τ_2	0.185 (0.010) [0.167, 0.204]	0.202 (0.010) [0.184, 0.220]	0.855
τ_3	0.227 (0.012) [0.206, 0.250]	0.208 (0.009) [0.190, 0.226]	0.292 (0.013) [0.266, 0.317]

注：τ = 潜在加工速度; $\mathbf{\Sigma}_\tau$ = 多维潜在加工速度的方差-协方差矩阵；上三角阵为相关系数，下三角阵为协方差；小括号内为标准误(后验分布标准差)；中括号内为 95%贝叶斯可信区间.

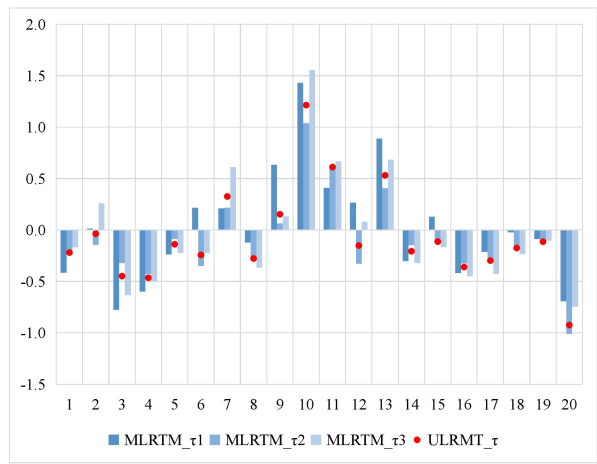


图 1. 2012 年 PISA 计算机化数学测验数据分析中前 20 名被试潜在加工速度估计值

注：ULRTM = 单维对数正态作答时间模型; MLRTM = 多维对数正态作答时间模型; τ = 潜在加工速度.

表 6. 2012 年 PISA 计算机化数学测验数据分析中题目参数估计值

题目	ULRTM						MLRTM					
	ξ			ω			ξ			ω		
	Mean	SE	95% CI	Mean	SE	95% CI	Mean	SE	95% CI	Mean	SE	95% CI
1	4.470	0.020	[4.432,4.508]	1.617	0.031	[1.558,1.678]	4.469	0.020	[4.433,4.510]	1.845	0.045	[1.760,1.936]
2	4.630	0.019	[4.592,4.667]	1.697	0.032	[1.635,1.762]	4.629	0.019	[4.594,4.668]	1.976	0.051	[1.874,2.076]
3	4.778	0.016	[4.750,4.811]	2.423	0.050	[2.327,2.519]	4.778	0.015	[4.747,4.807]	2.505	0.055	[2.397,2.612]
4	3.860	0.018	[3.825,3.895]	1.866	0.036	[1.793,1.934]	3.859	0.017	[3.825,3.894]	1.915	0.038	[1.841,1.991]
5	4.258	0.016	[4.226,4.291]	2.186	0.044	[2.104,2.274]	4.258	0.016	[4.224,4.287]	2.202	0.047	[2.112,2.295]
6	3.739	0.017	[3.707,3.774]	2.031	0.040	[1.958,2.116]	3.739	0.017	[3.706,3.771]	2.097	0.043	[2.012,2.179]
7	4.190	0.016	[4.158,4.220]	2.314	0.047	[2.221,2.406]	4.189	0.017	[4.156,4.222]	2.516	0.063	[2.393,2.638]
8	4.522	0.018	[4.487,4.557]	1.879	0.036	[1.809,1.950]	4.522	0.018	[4.488,4.558]	2.091	0.047	[1.995,2.180]
9	4.377	0.020	[4.338,4.417]	1.600	0.031	[1.533,1.656]	4.379	0.021	[4.339,4.420]	1.701	0.036	[1.632,1.771]

μ_ξ	4.316	0.202	[3.901,4.701]		4.315	0.199	[3.914,4.708]	
σ_ξ^2	0.367	0.217	[0.103,0.751]		0.366	0.219	[0.113,0.763]	

注：ULRTM = 单维对数正态作答时间模型; MLRTM = 多维对数正态作答时间模型; Mean = 后验均值; SE = 标准误(后验分布标准差); 95% CI = 95%贝叶斯可信区间.

4 模拟研究

上文已经通过实证研究阐述了 MLRTM 的实用性。进一步，我们使用两则模拟研究来探究新模型的心理计量学性能，以期进一步验证实证数据分析中所得到的结论。两个模拟研究均基于实证研究情境，其中，研究 1 拟探究(1) MLRTM 的参数估计返真性和(2)忽略潜在加工速度多维性所带来的影响。此时，使用 MLRTM 作为数据生成模型，并使用 MLRTM 和 ULRTM 进行参数估计。研究 2 拟探究冗余地指定潜在加工速度具有多维性所带来的影响。此时，使用 ULRTM 作为数据生成模型，并使用 MLRTM 和 ULRTM 进行参数估计。

4.1 模拟研究 1

4.1.1 数据生成与分析

模拟研究 1 中，设定 30 道题目考查 4 个维度，对应的 Q 矩阵呈现在图 2 中。参考实证研究中的估计值来设定模型参数的真值。对题目参数而言，时间强度参数依据$\xi_i \sim N(4, 0.25)$生成；而时间区分度参数依据$\omega_i \sim N(2, 0.25)$生成。被试量 $N = 1000$，多维潜在加工速度参数依据四元正态分布生成

$$\begin{pmatrix} \tau_{n1} \\ \tau_{n2} \\ \tau_{n3} \\ \tau_{n4} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.25 & & & \\ 0.15 & 0.25 & & \\ 0.15 & 0.15 & 0.25 & \\ 0.15 & 0.15 & 0.15 & 0.25 \end{pmatrix} \right),$$

该设定下，$\rho_{\tau\tau'} = 0.6$。基于 MLRTM 生成 50 组 RT 数据。

分别使用 MLRTM 和 ULRTM 去拟合生成数据。对于每组数据，马尔可夫链数、迭代数和预热数等均与实证研究中保持一致。采用 bias 和 RMSE 来评估参数估计返真性；另外，也计算了各参数估计值与其真值之间的相关系数(Cor)。

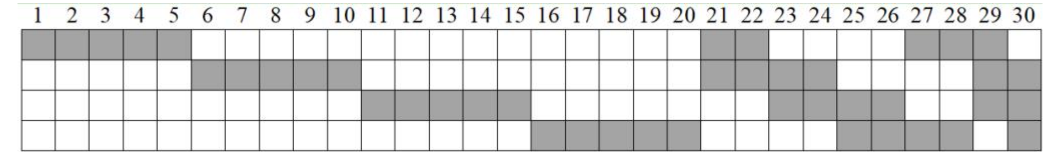


图 2. 模拟研究 1 中 $K \times I$ 的 Q'矩阵

注：灰色为 1、白色为 0.

4.1.2 结果

图 3 呈现了题目参数返真性。首先，整体来看 MLRTM 的返真性较好。其次，对时间强度参数而

言，两模型的返真性较为接近。对时间区分度参数而言，MLRTM 的返真性要优于 ULRTM 的返真性，尤其是对题目内多维题目。明确地说，对时间区分度参数而言，ULRTM 的 bias 和 RMSE 在题目间多维题目(题目 1 ~ 20)上分别约为–0.30 和 0.35；在题目内两维题目(题目 21 ~ 28)上约为–0.60 和 0.65；在题目内三维题目(题目 29 ~ 30)上约为–1.0 和 1.0。即 ULRTM 整体会低估题目区分度参数，这与实证数据分析中的结论相一致；此外，ULRTM 对题目区分度参数的返真性会随着题目所考查的维度数量增加而变差。

表 7 总结了被试参数的返真性。对每一个维度而言，所有被试的平均绝对 bias 和平均 RMSE 均分别约为 0.016 和 0.145，且所有被试的真值和估计值之间的相关系数也高于 0.95。表 8 呈现了被试参数方差-协方差矩阵的返真性。所有参数的 bias 和 RMSE 均接近于 0，返真性很好。

总之，根据模拟研究结果表明 MLRTM 可以得到较好的参数估计返真性。当数据包含潜在的多维潜在加工速度时，使用 ULRTM 会低估时间区分度参数，而时间强度参数几乎不受影响。

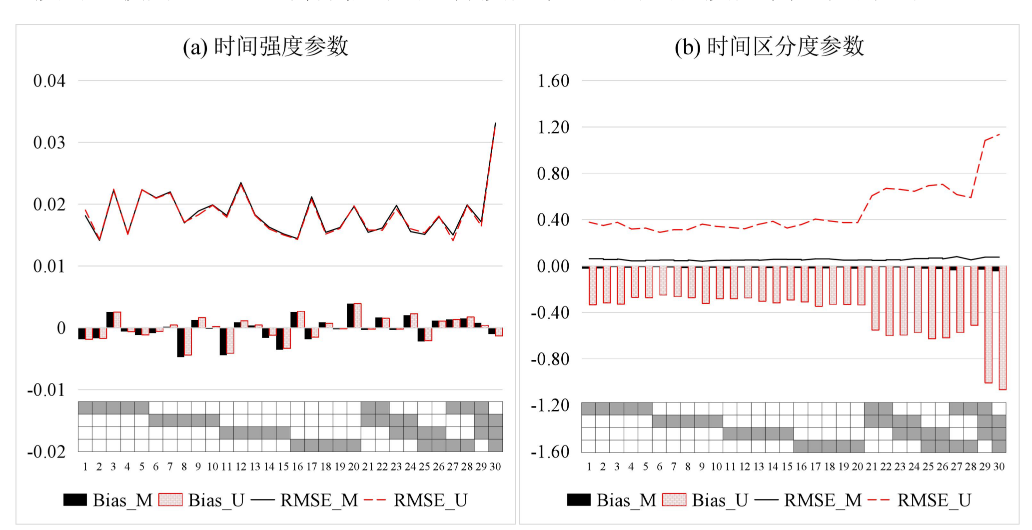


图 3. 模拟研究 1 中题目参数返真性(题目水平)

注：U = 单维对数正态作答时间模型; M = 多维对数正态作答时间模型; RMSE = 均方根误差.

表 7. 模拟研究 1 中被试参数返真性的总结

Parameter	MA_bias	M_RMSE	Cor
τ_1	0.016	0.147	0.956
τ_2	0.017	0.147	0.955
τ_3	0.016	0.144	0.957
τ_4	0.017	0.143	0.958

注：τ = 潜在加工速度; MA_bias = 所有被试的 bias 的绝对均值；M_RMSE = 所有被试的 RMSE 的均值; Cor = 所有被试的真值与估计值之间的相关系数.

表 8. 模拟研究 1 中被试参数的方差协方差矩阵返真性

Σ_τ	τ_1	τ_2	τ_3	τ_4
τ_1	0.00003 (0.00000)			
τ_2	0.00023 (0.00003)	0.00069 (–0.00010)		
τ_3	0.00031 (0.00004)	0.00015 (0.00002)	0.00015 (0.00002)	
τ_4	0.00015 (0.00002)	0.00041 (–0.00006)	0.00020 (0.00003)	0.00079 (–0.00011)

注：τ = 潜在加工速度; **Σ**$_τ$ = 多维潜在加工速度的方差-协方差矩阵；括号内为均方根误差(RMSE)；括号外为 bias.

4.2 模拟研究 2

4.2.1 数据生成与分析

模拟研究 2 中，设定 30 道题目考查单一维度。同样参考实证研究中的估计值来设定模型参数的真值。对题目参数而言，时间强度参数依据$\xi_i \sim N(4, 0.25)$生成；而时间区分度参数依据$\omega_i \sim N(2, 0.25)$生成。被试量 $N = 1000$，单维潜在加工速度参数依据$\tau_n \sim N(0, 0.25)$生成。基于 ULRTM 生成 50 组 RT 数据。同样，分别使用 MLRTM 和 ULRTM 去拟合生成数据；其中，使用 MLRTM 时冗余地将单维潜在结构设定为图 3 中的多维潜在结构。分析过程与指标等与模拟研究 1 保持一致。

4.2.2 结果

图 4 呈现了研究 2 中题目参数的返真性。对于题目时间强度参数而言，两模型的参数估计返真性基本一致。而对于题目时间区分度参数而言，MLRTM 的返真性略差于 ULRTM 的。再结合研究 1 中结果(见表 7)，发现冗余地指定潜在加工速度具有多维性所带来的负面影响低于忽略潜在加工速度多维性所带来的。

表 9 呈现了研究 2 中被试参数返真性。相比而言，MLRTM 的返真性略差于 ULRTM 的。但根据 Cor 指标可发现即便冗余地把单维结构指定为四个维度，每个维度的估计值与真值之间仍具有很高的相关系数。同时，我们计算了 MLRTM 中四个维度的潜在加工速度的估计值与 ULRTM 中单维潜在加工速度的估计值之间的相关系数，分别为$\rho_{\tau,\tau1} = 0.990$、$\rho_{\tau,\tau2} = 0.989$、$\rho_{\tau,\tau3} = 0.987$ 和$\rho_{\tau,\tau4} = 0.989$，即两模型的潜在加工速度估计值具有很高的一致性。此外，我们还计算了 MLRTM 中四个维度的潜在加工速度之间的相关系数，分别为$\rho_{\tau1,\tau2} = 0.979$、$\rho_{\tau1,\tau3} = 0.977$、$\rho_{\tau1,\tau4} = 0.981$、$\rho_{\tau2,\tau3} = 0.975$、$\rho_{\tau2,\tau4} = 0.978$ 和$\rho_{\tau3,\tau4} = 0.977$，即四个维度的估计值之间具有很高的相关性，表明它们测量/描述的很可能是同一个潜在变量。

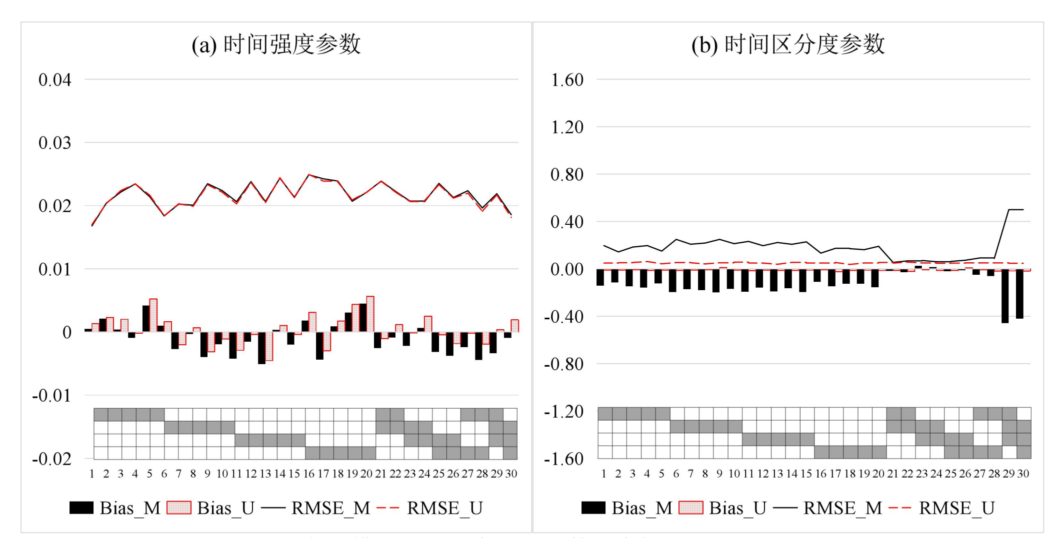


图 4. 模拟研究 2 中题目参数返真性(题目水平)

注：U = 单维对数正态作答时间模型; M = 多维对数正态作答时间模型; RMSE = 均方根误差.

表 9. 模拟研究 2 中被试参数返真性

分析模型	参数	MA_bias	M_RMSE	Cor
ULRTM	τ	0.013	0.088	0.985
MLRTM	τ_1	0.023	0.197	0.974
	τ_2	0.026	0.226	0.973
	τ_3	0.027	0.235	0.971
	τ_4	0.023	0.199	0.974

注：MLRTM 中各变量的返真性指标中真值均为单维潜在加工速度的生成值; τ = 潜在加工速度; MA_bias = 所有被试的 bias 的绝对均值；M_RMSE = 所有被试的 RMSE 的均值; Cor = 所有被试的真值与估计值之间的相关系数

5 总结与展望

为探究并分析多维测验中潜在加工速度的多维性，本研究提出了 MLRTM，新模型可视为对单维对数正态作答时间模型的多维拓广。随后，本文以 2012 年 PISA 计算机化数学测验中 RT 数据为例，通过探索性因素分析发现 RT 数据背后的多维潜在结构(即潜在加工速度的多维结构)与多维潜在能力的理论结构(即专家界定的 Q 矩阵)相匹配，验证了本研究的基本假设：在多维测验中，潜在加工速度具有与潜在能力相匹配的多维结构。然后，采用新模型对该数据做进一步分析，并与 ULRTM 的分析结果进行对比，结果表明在多维测验中考虑潜在加工速度的多维性是适合且必要的。最后，通过两则模拟研究探究了新模型的心理计量学性能，模拟研究 1 结果表明(1)贝叶斯 MCMC 算法能够为 MLRTM 提供较好的参数估计返真性；(2)忽略潜在加工速度的多维性对题目强度参数几乎无影响，但会大幅低估时间区分度参数，且返真性会随着题目所考查的维度数量增加而变差。模拟研究 2 结果表明(1)冗余地指定潜在加工速度具有多维性对题目强度参数几乎无影响，但会低估时间区分度参数；(2)当冗余地指定潜在加工速度具有多维性时，基于 MLRTM 的多维潜在加工速度估计值之间具有很高程度相关。此外，结合模拟研究 1 和 2 的结果，可发现(1)冗余地指定潜在加工速度具有多维性所带来的负面影响低于忽略其多维性所带来的；(2)当潜在加工速度具有多维潜在结构时(即 MLRTM 为数据生成模型)，使用 ULRTM 会低估时间区分度参数；而当潜在加工速度为单维结构时(即 ULRTM 为数据生成模型)，使用 MLRTM 也会低估时间区分度参数。因此，对时间区分度参数而言，当 ULRTM 的估计值小于 MLRTM 的时，可推断潜在加工速度具有多维结构；反之，当 ULRTM 的估计值大于 MLRTM 的时，可推断潜在加工速度具有单维结构。而实证研究中，ULRTM 对时间区分度参数的估计值小于 MLRTM 的，可推断实证研究中的潜在加工速度具有多维结构。

当然，尽管该研究得到了较好的结果，但由于能力和精力有限，本研究仍有一些局限性值得后续做进一步探究。首先，MLRTM 是对经典的 ULRTM 的多维扩展。由于对 RT 进行对数变换后仍有可能违反正态性假设，因此可尝试对本文所提出的 MLRTM 做进一步拓展，例如 Box-Cox 变换(Klein Entink et al., 2009)、线性变换(Wang, C. et al., 2013)以及 Log-Skew-Normal 变换(孟祥斌, 2016)等。其次，本研究提出的 MLRTM 为补偿模型，即假设多维潜在加工速度之间是相互补偿的。在题目内多维测验中，若

被试在某一维度中的潜在加工速度较慢，则可以通过在另一维度中的潜在加工速度来弥补。而至于潜在加工速度之间是否存在非补偿(或部分补偿)关系也值得今后做进一步探讨并开发相应的模型。再次，限于研究议题，本研究仅分析了 RT 数据，而没有同时对 RA 和 RT 数据进行联合分析。鉴于 RA 和 RT 数据同时包含被试和题目的信息，今后可基于贝叶斯层级建模框架，尝试建构可同时分析多维潜在能力和多维潜在加工速度的多维联合模型(Zhan, Jiao, Wang, & Man, 2018)；另外，MLRTM 是基于题目内多维度提出的，可同时处理题目内多维和题目间多维测验情境。但因为实证数据仅涉及题目间多维，所以从更严谨的角度看，实证研究结果仅为“潜在加工速度具有与潜在能力相匹配的*题目间多*维结构”提供证据。因此，尚缺乏证据表明“潜在加工速度具有与潜在能力相匹配的*题目内多*维结构”，有待后续研究进行补充。再另外，实证研究中的题目数量较少，可能会影响参数估计的精度和结论的准确性。因此，所得结论的普适性仍有待在更多的实证研究中进行验证。最后，本研究采用了相对简单的模拟研究来探究 MLRTM 的心理计量学性能，主要目的在于进一步支持实证研究中的结论。尽管研究结果表明新模型的参数估计返真性较好且为实证数据分析结果提供了支撑(例如，忽略潜在加工速度的多维性对题目时间强度参数无影响，但会低估题目时间区分度参数)，但未来仍可考虑增加模拟研究中的自变量(条件)，进而在更复杂、丰富的情境下探究新模型的心理计量学性能，为后续实证研究提供更丰富的理论参考。

参考文献

Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23.

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more from response times. *British Journal of Mathematical and Statistical Psychology, 71*, 13–38.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619.

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology, 10*:102.

Ferrando, P. J., & Lorenzo-Seva, U. (2007). A measurement model for Likert responses that incorporates response time. *Multivariate Behavioral Research, 42*, 675–706.

Fox, J.-P. & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement, 54*, 243–262.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. New York: Chapman & Hall.

Goudie, R. J., Turner, R. M., De Angelis, D., & Thomas, A. (2017). *MultiBUGS: A parallel implementation of the BUGS modelling framework for faster Bayesian inference*. arXiv preprint arXiv:1704.03216.

Guo, L. Shang, P., & Xia, L. (2017). Advantages and illustrations of application of response time model in psychological and educational testing. *Advances in Psychological Science, 25*, 701–712.

[郭磊, 尚鹏丽, 夏凌翔. (2017). 心理与教育测验中反应时模型应用的优势与举例. *心理科学进展, 25*, 701–712.]

Guo, X., Luo, Z., & Yu, X. (2020). A speed-accuracy tradeoff hierarchical model based on cognitive experiment. *Frontiers in Psychology, 10*:2910.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1–55.

Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in cognitive sciences, 3*(3), 91–98.

Lu, J., Wang, C., Zhang, J., & Tao, J. (2019). A mixture model for responses and response times with a higher-order ability structure to detect rapid guessing behaviour. *British Journal of Mathematical and Statistical Psychology*. Online First, https://doi.org/10.1111/bmsp.12175

Klein Entink, R. H., van der Linden, W. J., & Fox, J.-P. (2009). A Box-Cox normal model for response times. *British Journal of Mathematical and Statistical Psychology, 62*, 621–640.

Man, K., Harring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item responses and response times. *Applied Psychological Measurement, 43*, 639–654.

Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of neurology, 28*(5), 597–613.

Meng, X.-B. (2016). A log-skew-normal model for item response times. *Journal of Psychological Science, 39*, 727–734.

[孟祥斌. (2016). 项目反应时间的对数偏正态模型. *心理科学, 39*, 727–734.]

Muthén, L. K., & Muthén, B. (2019). *Mplus. The comprehensive modelling program for applied researchers: user's guide*, 5.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Manhattan: John Wiley & Sons.

OECD (2013). PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Reckase, M. (2009). *Multidimensional item response theory*. New York: Springer.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Tatsuoka, K. K. (1983). Rule Space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*, 345–354.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*, 181–204. http://dx.doi.org/10.3102/10769986031002181

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*, 287–308. http://dx.doi.org/10.1007/s11336-006-1478-z

van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement, 46*, 247–272. http://dx.doi.org/10.1111/j.1745-3984.2009.00080.x

van der Linden, W. J. (2011). Test design and speededness. *Journal of Educational Measurement*, 48, 44-60

van der Linden, W. J., Klein Entink, R., & Fox, J.-P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement, 34*, 327–347.

Wang, C., Chang, H., & Douglas, J. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology, 66*, 144–168.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology, 68*, 456–477.

Wang, C., Weiss, D. J., & Su, S. (2019). Modeling response time and responses in multidimensional health measurement. *Frontiers in Psychology, 10*:51.

Wang, S., Zhang, S., Douglas, J., & Culpepper, S. (2018). Using response times to assess learning progress: A joint model for responses and response times. *Measurement: Interdisciplinary Research and Perspectives, 16*, 45–58.

Wang, T., & Hanson, B. A. (2005). Development and calibration of an item response model that incorporates response time. *Applied Psychological Measurement, 29*, 323–339.

Zhan, P. (2019). Joint modeling for response times and response accuracy in computer-based multidimensional assessments. *Journal of Psychological Science, 42*, 170–178.

[詹沛达. (2019). 计算机化多维测验中作答时间和作答精度数据的联合分析. *心理科学, 42*, 170–178.]

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology, 71*, 262–286.

Zhan, P., Jiao, H., Wang, W.-C., and Man, K. (2018). *A multidimensional hierarchical framework for modeling speed and ability in computer-based multidimensional tests*. arXiv:1807.04003. Available online at: https://arxiv.org/abs/1807.04003

附录：MLRTM 中各待估计参数的先验分布设定

对于 MLRTM，首先，根据条件独立性假设，

$$\log T_{ni} \sim N(\xi_i - \sum_{k=1}^{K} q_{ik}\tau_{nk},\ \omega_i^{-2}).$$

其中，多维潜在加工速度向量的先验分布为：

$$\boldsymbol{\tau}_n \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\tau),$$

其中，方差-协方差矩阵的超先验为：

$$\boldsymbol{\Sigma}_\tau \sim \mathrm{InvWishart}(\mathbf{R},\ K),$$

其中，**R** 为 K 维对角矩阵。

对题目参数而言，

$$\xi_i \sim N(\mu_\xi, \sigma_\xi^2),$$

其中，均值和方差的超先验为：

$$\mu_\xi \sim N(4.3,\ 2) \text{ 和 } \sigma_\xi^2 \sim \mathrm{InvGamma}(1,\ 1).$$

Zhan 等(2018)的研究表明，对于 2012 年 PISA 计算机化数学测验数据中所有被试在所有题目上的平均 log RT 约为 4.301，因此我们将μ_ξ的均值设定 4.3。另外，$\omega_i^{-2} \sim \mathrm{InvGamma}(1,\ 1)$.

The multidimensional log-normal response time model: An exploration of the multidimensionality of latent processing speed

ZHAN Peida[1] Hong JIAO[2] Kaiwen MAN[3]

([1]Department of Psychology, College of Teacher Education, Zhejiang Normal University, Jinhua, 321004, *China*)

([2]Measurement, Statistics, and Evaluation, Department of Human Development and Quantitative Methodology, University of Maryland, College Park, Maryland, *United States*)

([3]Educational Studies in Psychology, Research Methodology, and Counseling, The University of Alabama, Tuscaloosa, *United States*)

Abstract

With the popularity of computer-based testings, the collection of item response times (RTs) and other process data has become a routine in large- and small-scale psychological and educational assessments. RTs not only provide information about the processing speed of respondents but also could be utilized to improve the measurement accuracy because the RTs are considered to convey a more synoptic depiction of the participants' performance beyond responses alone. In multidimensional assessments, various skills are often required to answer questions. The speed at which persons were applying a set of skills reflecting distinct cognitive dimensions could be considered as multidimensional as well. In other words, each latent ability was measured simultaneously with its corresponding working efficiency of applying a facet of skills in a multidimensional test. For example, the latent speed corresponding to the latent ability of decoding of an algebra question may differ from encoding. Therefore, a multidimensional RT model is needed to accommodate this scenario, which extends various currently proposed RT models assuming unidimensional processing speed.

To model the multidimensional structure of the latent processing speed, this study proposed a multidimensional log-normal response time model (MLRT) model, which is an extension of the unidimensional log-normal response time model (ULRTM) proposed by van der Linden (2006). Model parameters were estimated via the full Bayesian approach with the Markov chain Monte Carlo (MCMC). A PISA 2012 computer-based mathematics RT dataset was analyzed as a real data example. This dataset contains RTs of 1581 participants for 9 items. A Q-matrix (see Table 1) was prespecified based on the PISA 2012

mathematics assessment framework (see Zhan, Jiao, Liao, 2018); three dimensions were defined based on the mathematical content knowledge, which are: 1) change and relationships (θ_1), 2) space and shape (θ_2), and, 3) uncertainty and data (θ_3). One thing to note is that the defined Q-matrix served as a bridge to link items to the corresponding latent abilities, which shows the multidimensional structure of latent abilities. First, exploratory factor analysis (EFA) was conducted with the real dataset to manifest the multidimensional structure of the processing speed. Second, two RT models, i.e., the ULRTM and the MLRTM, were fitted to the data, and the results were compared. Third, a simulation study was conducted to evaluate the psychometric properties of the proposed model.

The results of the EFA indicated that the latent processing speed has a three-dimensional structure, which matches with the theoretical multidimensional structure of the latent abilities (i.e., the Q-matrix in Table 1). Furthermore, the ULRTM and the MLRTM yield adequate model data fits according to the posterior predictive model checking values (*ppp* = 0.597 for the ULRTM and *ppp* = 0.633 for the MLRTM). Furthermore, by comparing the values of the –2LL, DIC, and WAIC across the ULRTM and the MLRTM, the results indicate that the MLRTM fits the data better. In addition, the results show that (1) the correlations among three dimensions vary from medium to large (from 0.751 to 0.855); (2) the time-intensity parameters estimates of the two models were similar to each other. However, in terms of the time-discrimination parameters, the estimates of the ULRTM were slightly lower than the MLRTM. Moreover, the results from the simulation study show: 1) the model parameters were fully recovered with the Bayesian MCMC estimation algorithm; 2) the item time-discrimination parameter could be underestimated if the multidimensionality of the latent processing speed gets ignored, which meets our expectation, whereas the item time-intensity parameter stayed the same.

Overall, the proposed MLRTM performed well with the empirical data and was verified by the simulation study. In addition, the proposed model could facilitate practitioners in the use of the RT data to understand participants' complex behavioral characteristics.

Key words item response times; multidimensional latent processing speed; item response theory; computer-based testing; PISA